DOCUMENT RESUME

ED 344 928 TM 018 267

AUTHOR Hansen, Joe B.
TITLE A Purpose Driven Assessment Program.
PUB DATE Apr 92
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Criterion Referenced Tests; *Educational Assessment; Elementary Secondary Education; *Norm Referenced Tests; Portfolios (Background Materials); Public Schools; *School Districts; State Programs; Student Evaluation; Test Construction; *Testing Programs; Writing Evaluation
IDENTIFIERS *Colorado Springs Public Schools CO; Curriculum Based Assessment; Multiple Evaluations; Performance Based Evaluation; *Purpose Driven Assessment Program CO

ABSTRACT

The approach to assessment taken by the Colorado Springs (Colorado) Public Schools is outlined. Colorado Springs has taken a position similar to that advocated by the National Council on Educational Standards, that of multiple means of assessment applied in a dynamic and responsive manner. This approach in Colorado Springs is called Purpose Driven Assessment (PDA). Its primary component is an analytical framework based on an analysis of purpose patterned after R. Stiggins and N. Bridgeford (1987) and P. Airaisian (1984). The following three basic types of assessment are examined in relation to this model: (1) norm referenced testing; (2) goal referenced testing (criterion referenced); and (3) performance based assessment. Goal referenced testing in the Colorado Springs district is represented by a series of overlapping levels of tests, the District 11 Achievement Levels Tests (DALT), now used in grades 3 through 9. Performance based assessments are being developed, beginning with direct writing assessment, portfolio assessment, and curriculum-based measurement. Norm referenced testing in the Colorado Springs district is used to meet state testing requirements. Full implementation of the PDA model will require at least another 3 years, but the Colorado Springs school district is well on the road to matching assessment with its purposes. Four tables and two figures are included. There is a 10-item list of references. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED344928

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joe B. Hansen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A Purpose Driven Assessment Program

Joe B. Hansen, Ph.D.
Colorado Springs Public Schools

Presentation in the symposium:

MULTIDEMENSIONAL ASSESSMENT MODELS:
Matching Tools With Tasks

American Educational Research Association
San Francisco, April 23, 1992

TM018267

BEST COPY AVAILABLE

A Purpose Driven Assessment Program

Joe B. Hansen
Colorado Springs Public Schools

> Inherent in the general assessment standards is an assumption that all evaluation processes should use multiple assessment techniques that are aligned with the curriculum and consider the purpose of an assessment... (NCTM Standards, p.192)

INTRODUCTION

Never in the history of American education has so much attention been focused on student assessment as we have seen during the last few years. Since Cannell's (1989) scathing, though flawed, indictment of the evils of "high stakes" testing and the purported inadequacies of norm referenced tests, virtually everyone in the educational arena has joined the movement toward greater "authenticity" in the assessments we use, and toward providing a greater range of assessment options. This movement is occurring amidst a national resurgence of accountability and school reform. One result of these conditions is a convergence of the authentic assessment movement with the accountability movement. Nowhere is this more evident than in the efforts of the National Council on Educational Standards and Testing (NCEST).

An outgrowth of the President's Education Summit in Charlottesville, in 1989, the NCEST was formed by an act of Congress to:

- advise on the desirability and feasibility of establishing national standards and tests, and
- recommend long-term policies, structures and mechanisms for setting voluntary education standards and planning appropriate system of tests. (NCEST, 1992)

In its recently released report, NCEST endorses the establishment of high national standards for student performance and a system of assessments to measure performance against those standards. The NCEST recommended system of assessments should have two components - individual student assessments and a large scale sample assessment (such as) the National Assessment of Educational Progress. Further, the system of assessments should have the following features:

- multiple methods of measuring progress, not a single test,
- the system must be voluntary, not mandatory, and

- the system must be developmental, not static. (NCEST, p.4)

Regardless of the controversy surrounding the desirability and specific design of a national assessment system, the NCEST has taken a leadership role in advocating multiple methods of assessment, voluntarily applied in a dynamic and responsive manner. As I will illustrate, this position is similar to that which we have adopted in the Colorado Springs Public Schools.

Colorado State Mandate for Proficiencies Assessment.

Historically, the accountability movement has waxed and waned since the late 1800s, with recurring cycles of increased government involvement and private sector influence (Hansen, 1991). Colorado has had an educational accountability law since 1971. Rewritten in 1988, that law now embodies a set of goals developed by the State Board of Education. The goals address the areas of student achievement, graduation rate and attendance. One of those goals specifically requires a proficiency assessment system for every school district.

- By July 1, 1991, Colorado's school districts will have defined their expected student proficiencies at designated grade levels and will have implemented a program of measurement and reporting. (Colorado Department of Education, 1989)

These requirements are also incorporated into the State Accreditation Standards for school districts.

Elsewhere in the same law, the state requires the use of a standardized, norm-referenced test to assess the performance of all students in at least one grade level in each school, annually. The purpose of which is to facilitate comparisons between schools and districts. This is juxtaposed by language in the law that strongly encourages the use of alternative assessments "...such as portfolios, demonstrations, and performance assessments."

One could view this montage of assessment requirements as hopelessly confusing and impossible to deal with, or one could recognize the implicit, if not intentional, stimulus to develop a system of differentiated assessments. In Colorado Springs Public Schools we chose the latter. Our approach is based on the precept that different types of assessments must be used to address the many purposes for which assessment data are needed. No single assessment technique can serve multiple purposes equally well, therefore the development of a multidimensional approach to assessment is essential.

For our district and many others in Colorado this philosophy represents a significant break with tradition. Virtually all of Colorado's 176 districts have relied on multiple-choice, norm-referenced tests (NRT) to meet all of their formal assessment needs for many decades. Teachers, administrators and the public are accustomed to the annual or semi-annual ritual of preparing for, administering and receiving reports on, these tests. We have trained these user groups well in using these tests to make a myriad of educational decisions, from individual student placement to evaluating program efficacy. And they have learned well; too well it seems. Now, as we attempt to wean them away from over-use of and dependency on these NRTs, and orient them toward other measures, better suited for classroom decision making, we are faced with some degree of resistance. Many school district personnel, business people and parents believe strongly in these familiar tests and the data they produce, and they are loathe to let them go.

In recent years however, since the authentic assessment movement has been gathering momentum, many of these same teachers, administrators and parents have heard the call for greater authenticity and have flocked liked mice to the piper's tune, abandoning all reverence for the NRT they once held sacred. To illustrate this phenomenon, imagine education as a cruise ship, on which the passengers have been milling about, engaged in their various pastimes. Then, suddenly someone spots a large whale surfacing very near the boat and everyone runs to that side at the same time to catch a glimpse of the leviathan. This sudden shifting of weight on board the boat, from a random distribution to a concentration at one gunwale, could, if the boat lacks the appropriate combination of beam, ballast and keel, cause the craft to suddenly list to one side, spilling passengers overboard, and capsizing from the sudden shift of weight. Authentic assessments can in some respects be likened to the allegorical whale. If the stampede to one side of the boat is great enough the educational ship could capsize. To maintain a balanced ship we must avoid the headlong rush toward one form of assessment, to the exclusion of others, and learn to use a variety of assessment tools appropriately; using each for the purposes for which it is best suited.

Purpose Driven Assessment

In Colorado Springs Public Schools we call our approach to assessment - Purpose Driven Assessment or PDA, because we are consciously trying to balance the boat, based on the purpose for which assessment data are needed. Assessment tools are used for a wide variety of purposes and we believe strongly that the choice of tool should derive from the purpose for which the assessment data will be used. Numerous researchers have explored the issue of assessment purpose, and we have drawn from their work.

Research on Assessment Purposes

Several purposes for assessment have been identified through research. One taxonomy of purpose can be adapted from Stiggins and Bridgeford (1985) based on research on classroom assessments. These researchers defined classroom testing in terms of five purposes based on the types of decisions that teachers make: diagnosing, grouping, grading, evaluating, and reporting. Distinctions are also made across types of tests, e.g. publishers', teacher made, performance, etc. This has been extended as follows: Diagnosing strengths and weaknesses of individual pupils, diagnosing group needs, grouping students for instruction, determining the achievement potential of students, assigning grades to students, evaluating the instructional unit, communicating academic expectations, and controlling and motivating students. Other purpose factors identified by Stiggins (1987), affecting assessment, include identifying students for special services, communicating affective or behavioral expectations, and providing test taking experiences.

Airasian (1984) offers another taxonomy of purpose, in which two general levels of assessment are distinguished, 1.) externally mandated assessments, and 2.) teacher used assessments. The former includes high school graduation tests, preschool screening procedures, and tests to allocate resources, such as for a remedial reading program. These are typically mandated by an authority external to the school, such as the state school board or the legislature. To this list one could add statewide assessment programs and tests required by state accountability programs. These assessments serve the purpose of providing policy makers, data for monitoring the progress of schools and districts on state mandated goals or curriculum. Airasian's second level, those assessments used by teachers in their classrooms, are the assessments used to guide their instruction, maintain order, and to assign grades in their classrooms. These two levels of assessment have distinctively different characteristics.

Externally mandated assessments involve tests, regulations, objectives and passing scores set by the external authority. Their purpose is often external and may include a moral or managerial dimension. They rely heavily on a single test score and a single cutoff point for decision making.

Classroom assessments are performed on a day to day basis. They vary in form, e.g. standardized, non-standardized, teacher made (oral, observation, paper-pencil, essay or objective), performance checklists, rating forms, work samples, anecdotal records.

As the works of Stiggins and Airasian emphasize, the purpose for which assessment data are needed is a key determinant in

deciding what type of assessment to employ, in a given situation. Obviously, purposes vary across a broad spectrum and carry implications based on whether they are internally focused or externally imposed.

A PURPOSE DRIVEN ASSESSMENT MODEL

In our school district of 30,000 students, there are many purposes for assessment from screening and selecting students for special programs to summative program evaluation. Our model derives from an analysis of assessment purposes and a partitioning of all assessments into three basic categories: norm-referenced tests(NRT), goal referenced tests(GRT) and performance-based assessments(PBA.)

We have chosen to use the term "goal referenced" rather than the more traditional criterion referenced or CRT because CRT is often misused to refer to any objective based test, regardless of whether it is used in conjunction with performance criteria.

In its most rudimentary form our model is comprised of three boxes as shown in figure 1.

(Insert figure 1 here.)

The primary component of our PDA model is an analytical framework based on an analysis of purpose patterned after Stiggins and Bridgeford(1985), Stiggins(1987) and Arasian(1984). This component lists common purposes of assessment and examines each of the three basic types of assessments in the model in terms of their appropriateness for each purpose. This analytic framework is shown in table 1.

(Insert Table I, here.)

We acknowledge that our analysis of appropriateness of each type of assessment for the stated purpose is subject to dispute. Therefore I have designed a study to examine the validity of our analysis. That study will be the subject of a future paper or article and will therefore, not be discussed here. For the moment I ask that the reader accept our analysis as basically correct in the way it portrays the strengths and weaknesses of each type of assessment shown.

Another acknowledged flaw in this depiction is that PBAs come in many varieties and an analysis that accurately depicts one may not exactly fit another, depending on the specific characteristics of the assessment tool in question. For example, given the purpose of summative program evaluation and three types of PBA - portfolio assessment, direct writing assessment using an analytic trait model, and an oral presentation, it would be easier to obtain results that meet specified criteria of

validity, reliability, and aggregability using the analytic trait model than with either of the other two approaches. This is illustrated in Table II. Thus another dimension of analysis is needed, that of criteria for judging the quality of the PBAs. Both Arter(1991) and Linn, Baker and Dunbar(1991) have recently addressed this issue. Arter deals with "meta-rubrics" for determining the desired characteristics of a scoring rubric. Linn et al, on the other hand are concerned with appropriate

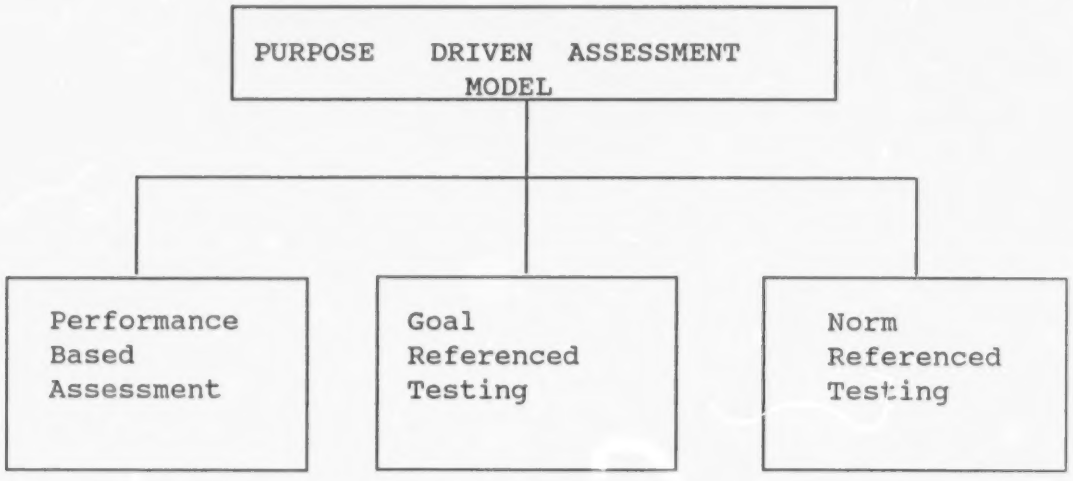

Figure 1. Purpose Driven Assessment Model

methods for determining the validity of performance assessments.

Obviously the issue of purpose of assessment is complex, and cannot be dealt with in a quick or superficial manner without substantial risk of oversimplification. Nonetheless, any analysis of assessment applications that begins with the question of purpose is likely to lead to more appropriate and effective uses of assessment than one that does not take purpose into account. Therefore we view our model not as the ultimate answer, but as a necessary foundation upon which we can eventually build a comprehensive assessment system.

Assessment techniques can also be evaluated in terms of how well they perform on the traits of information yield(Iy), validity(V), reliability(R), cost per pupil use(C), ease of use(E), and instructional value(Iv). My own analysis of the three major assessment types on these traits has led me to conclude that overall, GRT assessments have the best balance. This may be due, in part to the relative dearth of experience we have in the educational assessment world, with PBAs. Perhaps as we gain experience we will find satisfactory ways to improve reliability, verify validity assumptions, reduce costs - including time, while sustaining high information yields and

driving instruction toward mastery of more complex, integrated curriculum and higher cognitive levels.

Table I. Analysis of Assessment Type by Purpose of Assessment

Purpose of Assessment	Type of Assessment		
	NRT	GRT	PBA
Pupil selection and placement	x	x	x
Grouping for instruction		x	x
Diagnosis of instructional need		x	x
Evaluating achievement or growth			
- group	x	x	
- individual		x	x
Instructional program evaluation			
- program efficacy	x	x	
- identify strengths\weaknesses		x	x
Curriculum Evaluation			
- fidelity of instruction		x	x
- efficacy		x	x
- strengths and weaknesses		x	x
Certification of skills or knowledge		x	x
Geo-political comparisons with other schools, districts, states or nations		x	

Development of the Purpose Driven Assessment Model

One of the most pressing needs facing our district in 1989 was the proficiency assessment requirement of the state accreditation rules, borne out of the 1988 Accountability Act. The state accreditation rules require that

> By July 1, 1991, Colorado's school districts will have defined their expected student proficiencies at designated grade levels, and will have implemented a program of measurement and reporting. (Colorado Board of Education, 1989)

With a deadline looming only two years away in July of 1991, we knew we had little time to put a system in place for

proficiencies assessment. At that time the Iowa Test of Basic Skills (ITBS) was the mainstay of our testing program. We were very interested in beginning some pilot projects using alternative assessments such as portfolios, open-ended math problems and direct writing assessment, but we felt great time pressure to respond to the proficiencies assessment requirement. Our curriculum and instruction staff was developing a long range plan for defining proficiencies and examining graduation requirements in order to develop a certified diploma, which is also a state requirement that must be met by 1995.

Goal Referenced Testing

It was obvious that we needed to begin developing a proficiencies assessment system even before the proficiencies themselves were defined. We also recognized that we needed a system that would be dynamic and flexible in nature, so that it could be revised as the curriculum changed over time in response to societal pressures, community needs and educational theory. These needs, coupled with our analysis of purpose and type, led us to select a specific type of GRT as the centerpiece of our PDA model and to build other components around it. We chose a Rasch calibrated item-bank-based GRT system, developed and marketed by the Northwest Evaluation Association (NWEA). Using this system, which contains over 10,000 calibrated items in a computer managed item bank, we developed a series of overlapping levels of tests we call the District Eleven Achievement Levels Tests or DALT.

The DALT is a series of seven overlapping tests in each of two content areas, reading and mathematics, that measure the most significant outcomes, as determined by a group of master teachers, for grades three through nine. Because these tests follow a curricular continuum on which the degree of difficulty varies as you move to higher levels, one can, theoretically, predict whether a student's future performance will be at, above or below a predetermined standard. This is illustrated in figure 2. This is a highly desirable feature for a proficiency assessment in the basic skills area, since it allows for an early warning system for students in danger of not meeting the proficiency standards.

(Insert figure 2 here.)

The DALT system is now operational in grades three through nine. Even though the curriculum teams have not completed their work of defining proficiencies, we have an agreement with our colleagues in the Division of Instruction that the DALT will be the primary means of assessing proficiency. Over time, as exit outcomes are more clearly defined, performance assessments will be developed to complement the DALT and to bridge the gap from grade nine through twelve.

Table II. Analysis of strengths and weaknesses of three Performance Based Assessments as tools for summative program evaluation.

	Portfolio Assessment	Analytic Trait Writing Assessment	Demonstration by Oral Presentation
Strength	Provides rich detailed picture of student development. Drives instruction in a positive manner consistent with adopted curriculum. Encourages teacher-student interaction.	Provides quantitative data on multiple traits. Drives instruction in positive manner, based on writing process. Relatively high reliability, compared to other performance assessments. Easily aggregated data.	Corresponds more closely to real-life situation. Encourages integration of skills and knowledge across disciplines.
Weakness	Difficult to quantify or aggregate without compromising the holistic- qualitative concept of portfolio. Reliability questionable Can be very subjective if standards are not present and used. Reliability can be a problem. Validity is assumed, but not readily measurable.	Some educators may find it too reductionistic. Some may find it too structured or rigid, instructionally.	Inter-rater reliability is questionable unless a well defined scoring rubric exists and judges are well trained. Data can be difficult to aggregate. Predictive validity is assumed, but not known. Construct validity assumed, but not known.

11

12

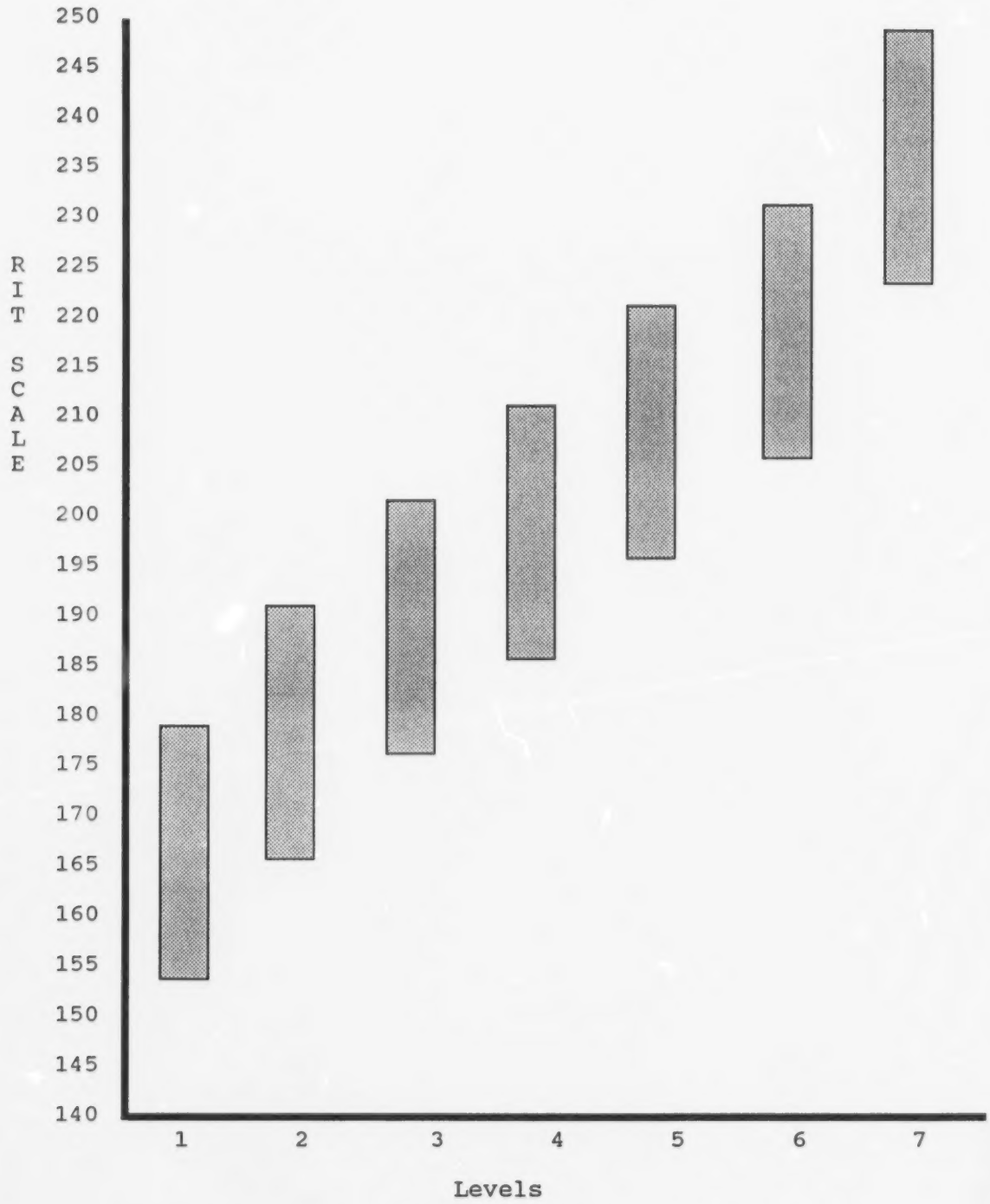

Figure 2. District Achievement Levels Tests

Purpose of the DALT. The DALT provides direct measures of knowledge acquisition and skill attainment on the goals upon which the curriculum is based. Examples of these goals are included in the appendix to this report. The DALT's fundamental purpose is to provide data on goal attainment by individual pupil, class, grade and school groups. It is used for thefollowing testing situations.

- Accountability pre-post testing for state mandated school improvement programs,
- Program and instructional effectiveness evaluation,
- Assessing student proficiencies at designated grade levels as required by state accountability law,
- Assessing exit outcomes for diploma certification, as required by state accountability law.

Performance Based Components

Under our framework, PBAs are most useful and effective at the classroom level, where the teacher, interacting with the student, can benefit from the greater depth and range of information they produce. As you move further from the assessment of individual pupil performance toward larger scale group assessments, a number of issues arise that pose difficulties for the use of certain PBAs, such as portfolios, in addition to the reliability and validity issues identified above. These issues include aggregation of data, and the managing and reporting of data. In my experience in working with teachers on the development of such assessments, they tend to have little concern for these problems since their focus is on helping the individual students. These issues have been discussed extensively elsewhere and will not be labored over here.[1] However we strongly advocate two key elements as essential components of PBAs, 1.) well defined scoring rubrics and 2.) careful training in the use of the rubrics. These are essential features for maintaining minimally acceptable levels of objectivity, assuring reliability and facilitating validity.

The PBAs we have begun to work with include direct writing assessment(DWA), portfolio assessment and curriculum based measurement(CBM.) Each of these assessment forms is identified with a particular assessment purpose or purposes.

---

[1]. See Frechtling, J. (1991). Performance Assessment: Moonstruck or the Real Thing? Educational Measurement Issues and Practice, 10(4), 23-25.

Direct Writing Assessment.

Direct writing assessment is a means of assessing student writing performance by obtaining writing samples produced in response to a common writing prompt, given to all students. The most common forms of DWA, based on the methods used for scoring the samples are holistic, primary trait and analytic trait.

Holistic scoring is based on the precepts that the whole is greater than the sum of its parts and that good writing can not be decomposed into specific and discrete components for the purposes of scoring. Instead, a piece of writing must be judged based on the total impression it gives the reader. Scoring may be done on a multi-point scale, but the entire piece is to be judged as a whole. Technical issues and mechanics such as spelling and punctuation are taken into account as they affect the overall impression of the rater. Raters are trained using anchor papers that represent points on the scale, e.g this paper is a "1", this is a "3" and so on.

Primary trait scoring is focused on the single most important trait or characteristic of the paper. For example the purpose of the paper might be to persuade someone else to vote for your candidate. The persuasiveness of the paper is therefore the primary trait. Scoring is done in a similar fashion to the holistic approach, but raters may be instructed not to rate on mechanical and technical errors such as punctuation, spelling or even minor grammatical errors.

Analytic-trait DWA is multidimensional. A paper may be rated on several traits. Analytic-trait models may use as few as four or as many as ten traits. Each paper is scored separately for each of the traits, by at least two raters. In cases of wide disparity between the two raters a third rater is usually asked to rate the paper.

The particular DWA model we have chosen is a six trait analytical model, in which the traits are

- Ideas / Content
- Organization
- Voice
- Word choice
- Sentence fluency
- Writing conventions

The model also embodies four writing modes: descriptive, persuasive, expository, and imaginative. Training and information on the use of this model is available through the Northwest Evaluation Association. It has been in use in many school districts in the Pacific Northwest for more than a decade.

In the late summer of 1991 we sent three testing and evaluation administrators, two instructional administrators and four teachers from our district to a three day, intensive training session on the six-trait analytic writing assessment model. This cross-sectional group formed a nucleus for training others in the district. In the early spring of 1992, twenty five more teachers were trained. We plan to expand the training in the summer of 1992. Our goal is to have the six-trait analytical model in use in each of our fifty three schools by the end of 1993-94 academic year. At that time we will begin collecting system-wide data on the use of the model.

Portfolio Assessment.

Not surprisingly, the portfolio assessment model we have chosen to work with is also based on a definition developed by NWEA.

> A portfolio is a purposeful collection of student work that exhibits the student's efforts, progress and achievements in one or more areas. The collection must include student participation in selecting contents, the criteria for selection, the criteria for judging merit, and evidence of student self-reflection. (Paulson, Paulson and Meyer, 1991.)

This definition grew out of a series of working retreats sponsored by NWEA in May and August of 1990, and March of 1991.

At the present time our portfolio efforts are limited to one elementary school to which we are providing direct support, and an unknown number of spontaneous but not well defined portfolio activities scattered throughout the district. One of our goals for 1992-93 is to conduct a portfolio assessment institute during the summer months for teachers wishing to participate in a district wide portfolio assessment pilot project.

Curriculum Based Measurement.

Curriculum based measurement is a very specific and limited form of PBA that uses text material extracted from the curriculum to obtain brief fluency measures of student mastery. In reading, it uses number of words read correctly in a timed, short oral reading exercise to ascertain student progress. Since CBM will be discussed in the presentation by Mengel and Schorr I will not

discuss it further here, except to say that we have assigned it a limited range of use based on purpose, for screening of special education and compensatory education students and for use by teachers of these students to obtain quick measures of student progress in basic skills mastery.

CBM has been subjected to criticism by those who feel that alternative assessments should address higher order cognitive skills. Higher order skills are a critical issue for assessment and we feel strongly that no system is complete unless it addresses this issue in a substantive way. Nonetheless, we feel equally strongly, that there is a need for the type of short fluency measure offered by CBM. Therefore we have included it in our PDA model.

Norm Referenced Testing

Current state accreditation regulations require the annual testing of all students in at least one grade in each school, using a norm-referenced test, in order to facilitate comparisons between schools and districts within the state and between Colorado students and the national norming population. Although we object to the use of test data in this political manner, we also recognize it as a commonly perceived need. Therefore, we have chosen to comply with the state's requirement at the minimally acceptable level, by allowing each principal to choose which grade to test each year. We have commissioned an equivalency study of the DALT and the ITBS and are hopeful that we will be able to obtain acceptably accurate NCE estimates using the DALT. If so, we will petition the state for a waiver to allow us to use the DALT instead of the ITBS. Messages from the state are mixed. New proposed accreditation regulations encourage multiple assessments and do not mention NRTs specifically, but they still call for comparisons across schools and districts.

We are currently administering the ITBS to grades 5, and 8 and the Test of Achievement and Proficiency (TAP) to grade 11 to comply with state regulations and to obtain an independent indicator of the general achievement levels of our students against a common standard, the publisher's norms. We recognize that our GRT and the PBAs we will be using increasingly in the future are not well understood by legislators, business people and the lay public, and therefore it may take several years before they are recognized as legitimate indicators of student skills and knowledge by those groups. Meanwhile we will continue to develop assessments, train teachers and educate our parents and public on the uses and limitations of various assessment types.

Table III contains a summary of the purposes for which we assessment and the particular types of assessments used.

Table III. Summary of Purposes for Assessment and Corresponding Types of Assessments Used in the PDA Model

Assessment Purpose	PDA Assessment
Assessing student proficiencies in mathematics, reading and language arts.	District Achievement Levels Tests, Grades 3 through 9. Portfolio Assessment - all grades, Performance Based Assessments to be developed Grades 10 -12.
Assessing student proficiencies in general science, social sciences.	Performance assessments to be developed.
Evaluating student mastery of course level subject matter for grades 10-12 in mathematics, humanities, science and social sciences.	DALT course level tests and Performance assessments to be developed.
Selection and placement of students for special education and compensatory education, grades 2-9.	Curriculum Based Measurement
Monitoring progress of special education and remedial students on mathematics, reading and writing, grades 2-9.	Curriculum Based Measurement, portfolio assessment.
Assessing student growth across subjects, primary (k-3) level.	Portfolio assessment
Assessing student growth in writing, grades 3-9.	Direct Writing Assessment - Six Trait Analytical Model
Evaluating language arts curriculum, grades 3-9.	DWA - Six Trait Model
Certified diploma, assessing student attainment of exit outcomes.	Written expression: DWA-Six Trait Model, Mathematics: DALT, Specific course content in mathematics, science, social sciences: DALT type course level tests to be developed. Employability and literacy skills: Portfolio system to be developed for grades 9-12.
Meeting state and federal requirements for Comparing performance of schools & programs against publisher norms	Norm Referenced Tests

Current Status and Planned Future Development

The PDA model, conceived of in 1989, is now in early stages of development and implementation. We expect that full implementation will require at least another three years, through the 1995-96, academic year.

Table IV. shows milestone events during each of the next three years.

(Insert Table IV here.)

Numerous obstacles must be surmounted if we are to succeed in realizing full implementation of our PDA model. To begin with, we face a serious funding crisis in our district, as do many other school districts in Colorado. State funding for education faces a $168 million shortfall in fiscal 1992 and the problem increases to $433 million in 1993, if it is not solved before then. This is a general election year and the Colorado state legislature has thus far not shown the courage to pass a bill that would generate the revenue needed to rescue the public school system. Our Governor, Roy Romer has threatened to veto any school finance bill that does not at least hold the current level of funding constant, thereby setting the stage for a possible showdown over school funding at the polls in November. We may find that we not only can't fund our PDA model, but that we will have to cut back on the staff needed to operate and maintain our DALT program. Conceivably, our entire department could be eliminated by next fall.

Assuming that we survive at a funding level that will permit us to pursue our plans, we have a herculean job of planning, training, field testing, evaluating and then repeating the cycle over the next several years, if we are to succeed.

With great amounts of luck and hard work, we hope to reach our goal of having the PDA fully implemented by 1996. Meanwhile, as computer hardware advances occur and current state-of-the-art equipment becomes more affordable, we hope, sometime in the near future to begin experimenting with computerized adaptive testing, using the NWEA system.

As this report is being written, the Colorado State Board of Education is conducting hearings on new proposed accreditation and accountability rules that will mandate the use of multiple measures for assessing student proficiencies, and certifying the diploma. Competing sets of achievement and assessment standards being considered by the State Board and legislature also specify multiple assessment techniques. It is encouraging to see that the bureaucrats are beginning to catch up with the state of the art as they develop policies and standards.

Table IV. Milestone Events in Continuing the Development and Implementation of the Purpose Driven Assessment Model in Colorado Springs Public Schools.

Time Period	Event
June 1992.....	Exit outcomes and graduation requirements defined.
June - July 1992.....	Develop DALT for Language Arts
August 1992.....	Summer Assessment Institute for teachers and principals • Purposes of assessment • Direct Writing Assessment Training • Portfolio Assessment training
September - October 1992.....	Field test and revise Language Arts DALT
	Establish portfolio assessment pilot program in 5 elementaries, 2 Junior Highs and 1 High School
	Establish district-wide pilot program on Direct Writing Assessment
January - June 1993.....	Identify course level DALT needs
June 1993.....	Begin course level DALT development in one subject area
	Evaluate portfolio assessment pilot progam
	Evaluate Direct Writing Assessment pilot program
August 1993.....	Assessment Training Institute for Teachers and Principals
September 1993.....	Expand Portfolio and DWA pilots to 10 elementaries, 4 Jr. Highs and 2 High Schools
September-October 1993.....	Field test course level DALT - course #1
June 1994.....	Evaluate Pilot Programs
	Begin developing course level DALT #2
August 1994.....	Continue Assessment Institute
September 1994.....	Expand DWA to all elementary and Jr. Highs and PA pilot programs to 20 elementaries, 6 Jr. Highs and 3 High Schools
September-October 1994.....	Field test course level DALT #2
June 1995.....	Evaluate pilot programs
	Begin development of course level DALT #3
August 1995.....	Continue assessment training institute
	Expand Districtwide pilot programs to all remaining schools
September-October 1995.....	Field test course level DALT #3
June 1996.....	Evaluate pilot programs
	Begin development of course level DALT #4

REFERENCES

Airasian, P. (1984). Classroom Assessment and Educational Improvement. Keynote address for conference: Classroom Assessment: a key to educational excellence. Portland, OR., Northwest Regional Educational Laboratory.

Arter, J.(1990). Performance Rubric Evaluation Form (metarubric). Portland, OR., Northwest Regional Educational Laboratory,

Cannell, J. (1989). How Public Educators Cheat on Standardized Tests. Albuquerque, NM. Friends for Education.

Colorado Department of Education (1989). 1989 Colorado State Board of Education Goals - Strategies for Excellence. Denver, CO. Colorado Dept. of Education.

Hansen, J. (1991) Accountability as a Tool for Educational Reform: Is it an Oxymoron? Presentation in the symposium: ACCOUNTABILITY LEGISLATION AND SCHOOL REFORM, Issues and Perspectives from Districts and States, at the conference of the American Educational Research Association, Chicago, April.

Linn, R., Baker, E. and Dunbar, S. (1991). Complex Performance-Based Assessment: Expectations and Validation Criteria. Educational Researcher, 20,(8), 15-21.

National Council on Educational Standards and Testing, (1992). Raising Standards for American Education. U.S. Govt. Printing Office, Washington, D.C.

Paulson, F., Paulson, P. and Meyer, C. (1991). What Makes a Portfolio a Portfolio? Educational Leadership, February, 60-63.

Stiggins, R. (1987). Profiling Classroom Assessment Environments. Paper presented at annual meeting of National Council on Measurement in Education, Washington, D.C.

Stiggins, R. and Bridgeford, N. (1985). The Ecology of Classroom Assessment. Journal of Educational Measurement, 22(4), Winter, 271-286.